

ジャパンリンクセンター

「研究データへのDOI登録実験プロジェクト」中間報告会 (2015. 07. 03)

---

千 葉 大 学

千葉大学附属図書館  
三 角 太 郎

## 実験には図書館として参加

---

大学図書館の使命は、**学術情報の流通**です。論文だけでなく、研究データも含めた多様なコンテンツに対応していきたいと考えています。

図書館として何ができるのか？それを探るために、実験に参加させていただきました。

# 1. テスト実施方針

---

研究データ管理のプラットフォームは、基本的には附属図書館で運用している**機関リポジトリシステム**を想定している。本テストでは、機関リポジトリシステムを研究データマネジメントのプラットフォームとして使用する際の課題点を抽出、解決策を検討し、最終的には実運用の業務フローの確立を目指す。本学では既存の研究データのデータベースをもつわけではなく、**研究成果のオープンアクセス化**の対象の一部として、研究データをとらえている。そのためDOIの登録テストについても、研究データに限定せず、論文等のDOI付与のフローもあわせて検討する。特に1)登録ポリシー 2)運用フロー 3)メタデータスキーマについての検討を行う。

# 機関リポジトリの運用指針

## 【運用指針の例】

### 千葉大学学術成果リポジトリ運用指針

平成 17 年 2 月 1 日  
附属図書館運営委員会制定

平成 22 年 3 月 30 日  
情報化推進企画室図書館専門部会改正

(千葉大学学術成果リポジトリ)

1. 千葉大学附属図書館は、千葉大学(以下「本学」という。)において作成された電子的な学術研究成果を収集し、千葉大学学術成果リポジトリ(以下「リポジトリ」という。)に恒久的に蓄積・保存し、学内外に無償で発信・提供することにより、**本学の学術研究の発展に資するとともに、社会に対する貢献**を果たすものとする。

(登録)

2. (登録対象となる学術研究成果は以下の要件を満たすものとする。

(1) **学術的な研究の成果**であること。

(2) 本学においてその主要な部分が作成されたものであること

(3) 電子的フォーマットで作成されていること

(4) ネットワークを通じて配信できること

## 2. テスト登録対象データと登録対象件数

---

本テストで計画したテスト登録は以下の二つのタイプ。

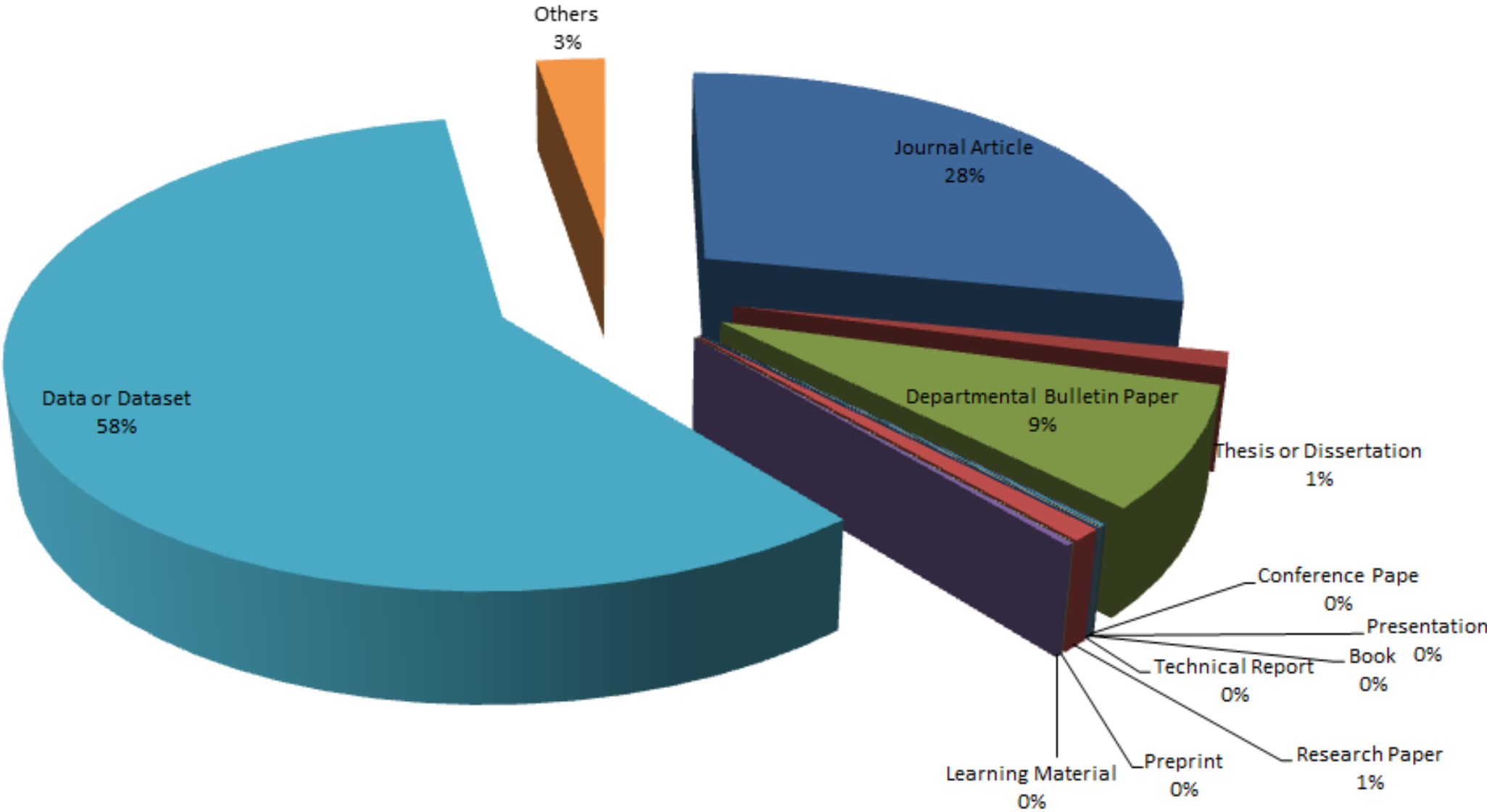
- 1) 既存の機関リポジトリシステム搭載のデータセット
- 2) 研究プロジェクトの成果物の登録シミュレーション

## 1) 既存の機関リポジトリシステム搭載のデータセット

---

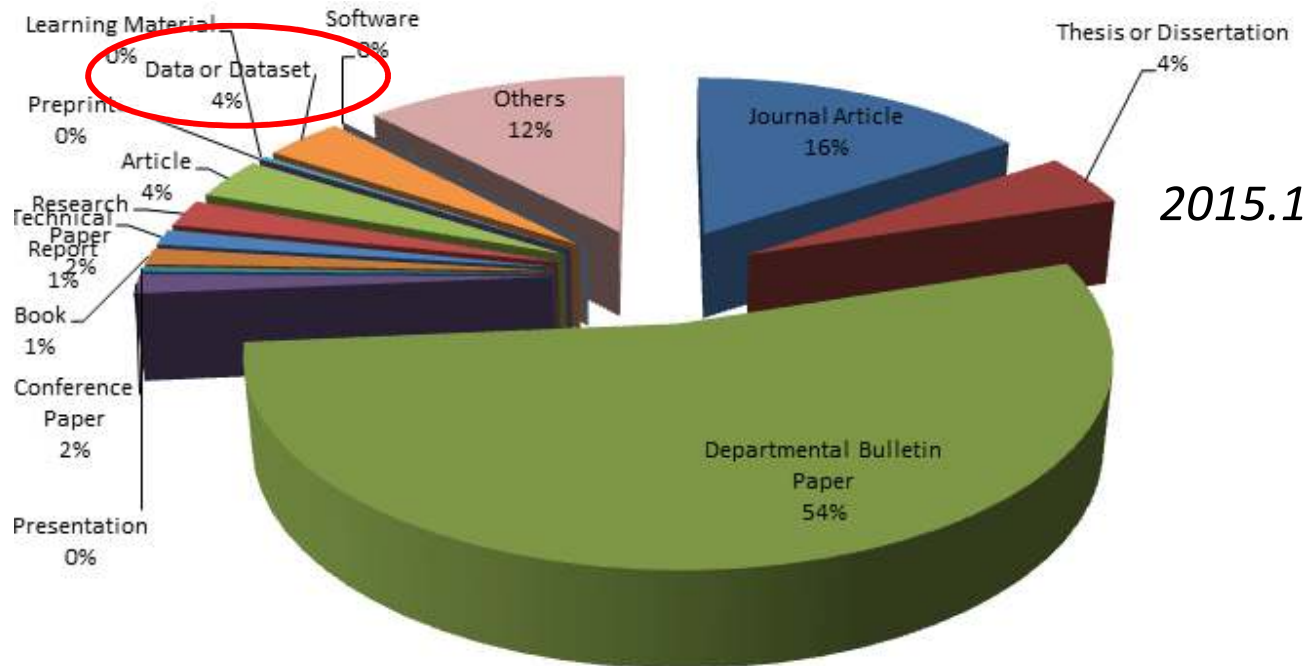
現在、運用中の機関リポジトリシステム搭載のデータのうち(表1)、資料種類がデータセットの約5万2千レコードのうちの代表的なもの100件程度へのDOI付与を行い、データ登録へ要する業務量の見積りも行う。

# 千葉大学機関リポジトリのコンテンツ構成



2015.1

# あらためて機関リポジトリの現状はどうなっているのか



日本の機関リポジトリ中のデータ

52,421件

そのうち千葉大学のコンテンツ

52,223件

**99.6%**



この画像、あなたなら、どう検索しますか？

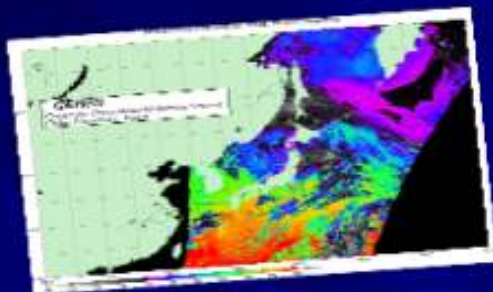


"Data-driven"

データの使い方を利用者に委ねることで研究の可能性を無限に拡げてゆくのがe-Scienceの魅力。

でも、アクセスできなければ、その可能性は眠ったまま。

「最良の検索キーは、利用者が知っている」これが私たちの出発点です。



Folksonomy が、e-Science を変える。

Social Tag が、データに命を与える。

「みんな」で築く e-Science、リポジトリから始まります。

教員と図書館のコラボレーション

2008年、IRコンテンツにソーシャルタグを付与するシステムを開発。



「メタデータから漏れていたキーワードが、タグによって補われていた」

「他のタグからの連想で、新たなタグが付与された」  
実証実験によって確認された効果の一例です。

2009年は「教員と図書館の連携」を図りながら、実証実験を拡大します。

e-Science基盤構築のためのデータ・キュレーション機能拡充の実証実験

代表機関：千葉大学 / 連携機関：北海道大学、金沢大学、大阪大学、九州大学

# e-Science Project

## CSI事業

### 2008～2009年度

実証実験として、コンテンツの作成者側と利用者側のそれぞれの立場から、e-Scienceデータ（主として千葉大学リモートセンシング研究センターの衛星画像）へ試行的にメタデータを付与できる環境を用意し、そのメタデータの評価をとおして、異種データの格納が可能なユーザコミュニティの需要に応えるメタデータ形式の検討を行い、人類の知的生産物の管理ツールとして機関リポジトリが位置づけられることをめざす。

# 菘庭標本



## 2) 研究プロジェクトの成果物の登録シミュレーション

---

現在、本学のアカデミック・リンク・センターで実施されている研究プロジェクトの成果である研究データについて、機関リポジトリ上でのデータ保存・公開・DOI付与のシミュレーションを行う。本テスト内で、実際にDOIを付与するかどうかは、シミュレーションの検討結果によって決定する。アカデミック・リンク・センターのプロジェクトをとりあげるのは附属図書館サイドでハンドリングが容易なためである。アカデミック・リンク・センターは、アクティブ・ラーニングについての実践的な研究をミッションとしているが、本テストの実施母体である附属図書館と連携して研究を行っているため、プロジェクト研究活動の全体像を附属図書館サイドでも把握しているためである。

## 2) 研究プロジェクトの成果物の登録シミュレーション

---

実際の登録時には、大学内で進められている研究プロジェクトについて、学内の関連部署とも連携しながら、成果の登録・保存・公開の業務を進めることになるが、本テストではプロジェクトのスタートから終了までの研究サイクルの各フェーズについて、実際にどのような業務を行うべきかを検討し、課題の抽出をはかる。研究データ以外の生産物については、本テストの対象外であり、本テスト内ではDOIは付与しないが、業務フローとしては不可分のものであると予想されるため、検討の対象として含める。具体的に想定している研究データとしては、画像、動画、数値データを想定し、その他に公表論文、講演資料、各種報告書を想定している。

# 実験結果

---

## (i)登録インターフェース

当初は、WebとXMLの双方をテストする計画であったがWebのみの個別登録にとどまっている。本登録時には、数千件オーダーのデータアップロードが必要であるが、今回登録時にEXCELにて作成したメタデータについて、インポートに失敗。データを視認した限りでは不備は見えないため、改行コード、区切記号等が原因ではないかと考えている。その原因について解決できていない。しかし実験では、メタデータ仕様等に検討を特化することとしたため、XML登録については、本登録時までテストを先送りすることとした。

# 実験結果

---

## (ii) テスト登録

コンテンツについて、今回は本学の機関リポジトリシステムのデータセット中の最大のコレクションである、萩庭コレクション(植物標本)にフォーカスを絞ることとし、複数件について登録テストを実施した。他の学術雑誌掲載論文の附属データについて、昨年度より論文本部の登録要請時に、附属データについても登録可否をあわせて問い合わせをはじめたが、現時点ではほとんど登録が進んでいない。

# 実験結果

---

植物標本データについて、採集者/標本作成者/デジタルファイル作成者等の複数の人間が作成にかかわっているが、それをメタデータへどのように記述するかで苦慮した。実験データごとに、メタデータの記述方法に大きなバリエーションはあることが予想されるが、**最低限の入力ガイドライン**は必要。

# 実験結果

---

研究プロジェクトの成果物の登録シミュレーションについては、検証が十分にできなかった。本センターで扱うデータは、基本的に利用者の個人情報絡む。論文、報告書等にもちいる場合には抽象化・統計化し、データの解析・保存はスタンドアローンのPCで行うなどの、十分な注意をはらっている。しかし**研究データとして保存する場合に、どの段階のデータを残すべきか**は、まだまだ議論が必要。おそらく社会科学系のデータについては同様の問題が生じるが、調査結果を統計的に分析して公表する場合に、分析した元データが個人情報を十分に抽象化してから統計分析を行っているのか、統計分析を行った結果として個人情報が抽象化されたのであって、元データは個人情報が特定できる状態なのか？



# あらためて図書館の特徴を考えて見る

---

## ○ 学術情報の**アーカイブ**をミッション

→ 大学内でアーカイブの文化をもつ組織は図書館・博物館・記念室ぐらい。

## ○ 書誌情報(メタデータ)作成のノウハウ

→ まともな図書館員ならある程度の書誌作成のスキルはもっている

## ○ 中身を**理解できなくても扱**うことができる

→ まったく**理解できない**専門的な数学書の書誌を作ったり、**読めない**外国語の書籍の書誌を作ったりを**業務としてこなす**ようトレーニングされている

# あらためて、図書館から どうアプローチするかを考えてみる

---

## 1) 最初に研究ありき

→ 最終的な研究成果物である論文が  
あって、その添付物としての研究データ  
がある

## 2) 最初にデータありき

→ 博物館標本の画像・計測データとか  
観測データとか、まずはデータがあって、  
それがみな論文になっているとは  
限らない

# アプローチ1

---

## 1) 最初に研究ありき

→ 最終的な研究成果物である論文が  
あって、その添付物としての研究データ  
がある

図書館員にはわかりやすいアプローチだが、しかし、  
いつまで論文が最終的な研究成果物であるかもわから  
ない。研究データ流通の構造そのものが大きく変わ  
ろうとしているのでは？

論文は研究データのメタデータ？

## アプローチ2

---

### 2) 最初にデータありき

→ 博物館標本の画像・計測データとか  
観測データとか、まずはデータが  
あって、それがみな論文になっている  
とは限らない

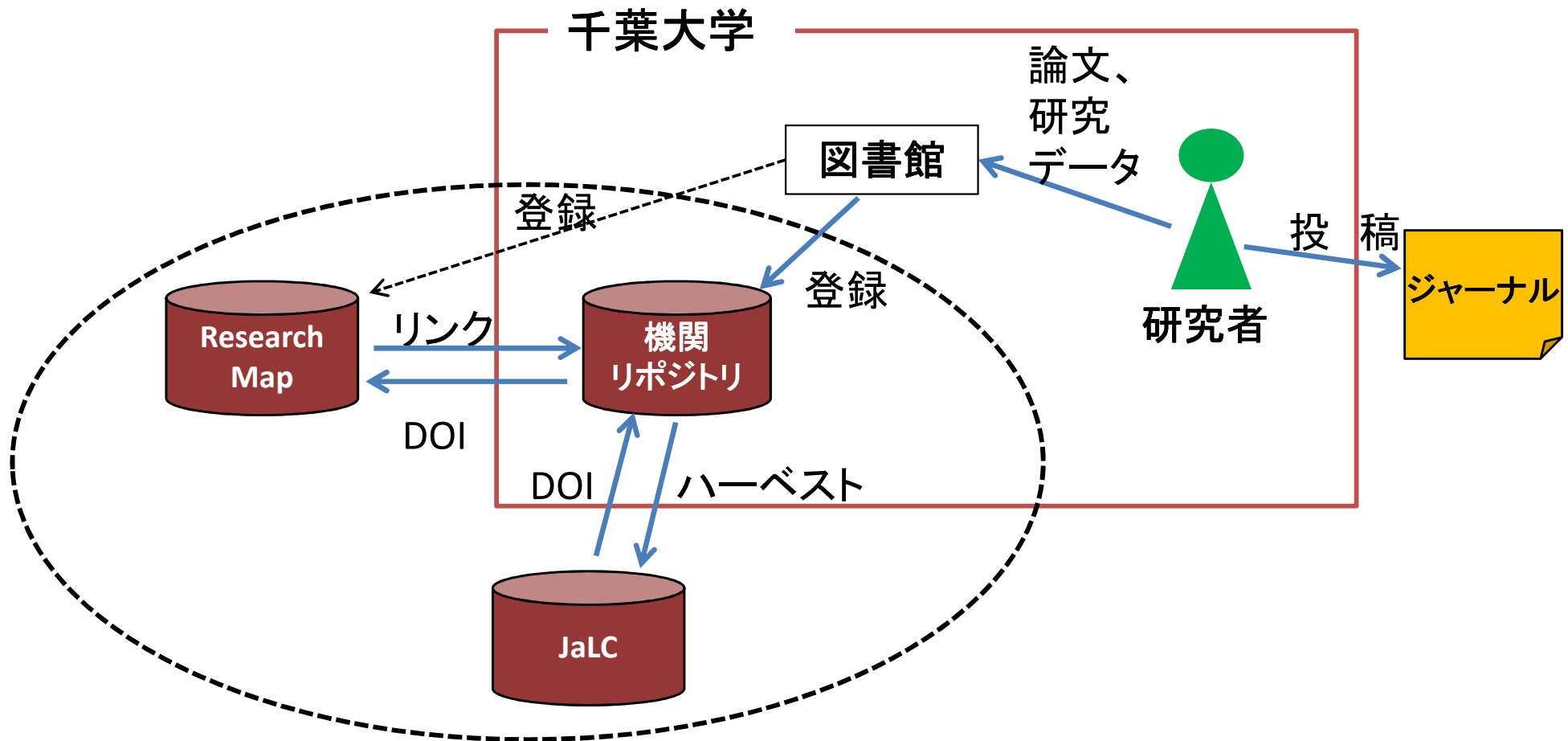
テキスト情報が付与されていないもののメタデータ  
は困難。

ノウハウがあれば、ある程度は・・・。

# 課題：業務フロー検討案

## 1)一般的な研究データ登録フロー

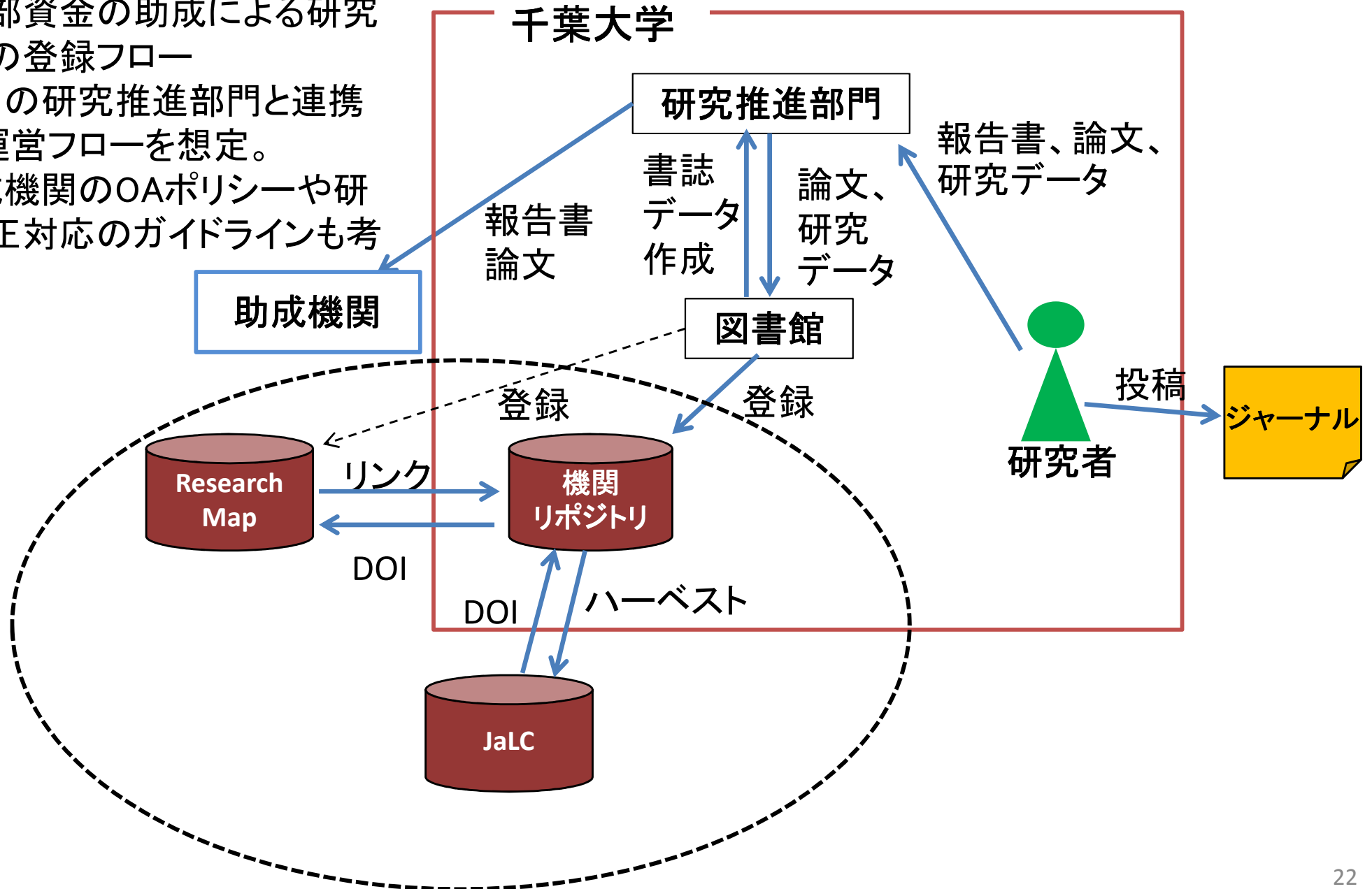
図書館と研究者と連携した運営フローを想定。



# 課題：業務フロー検討案

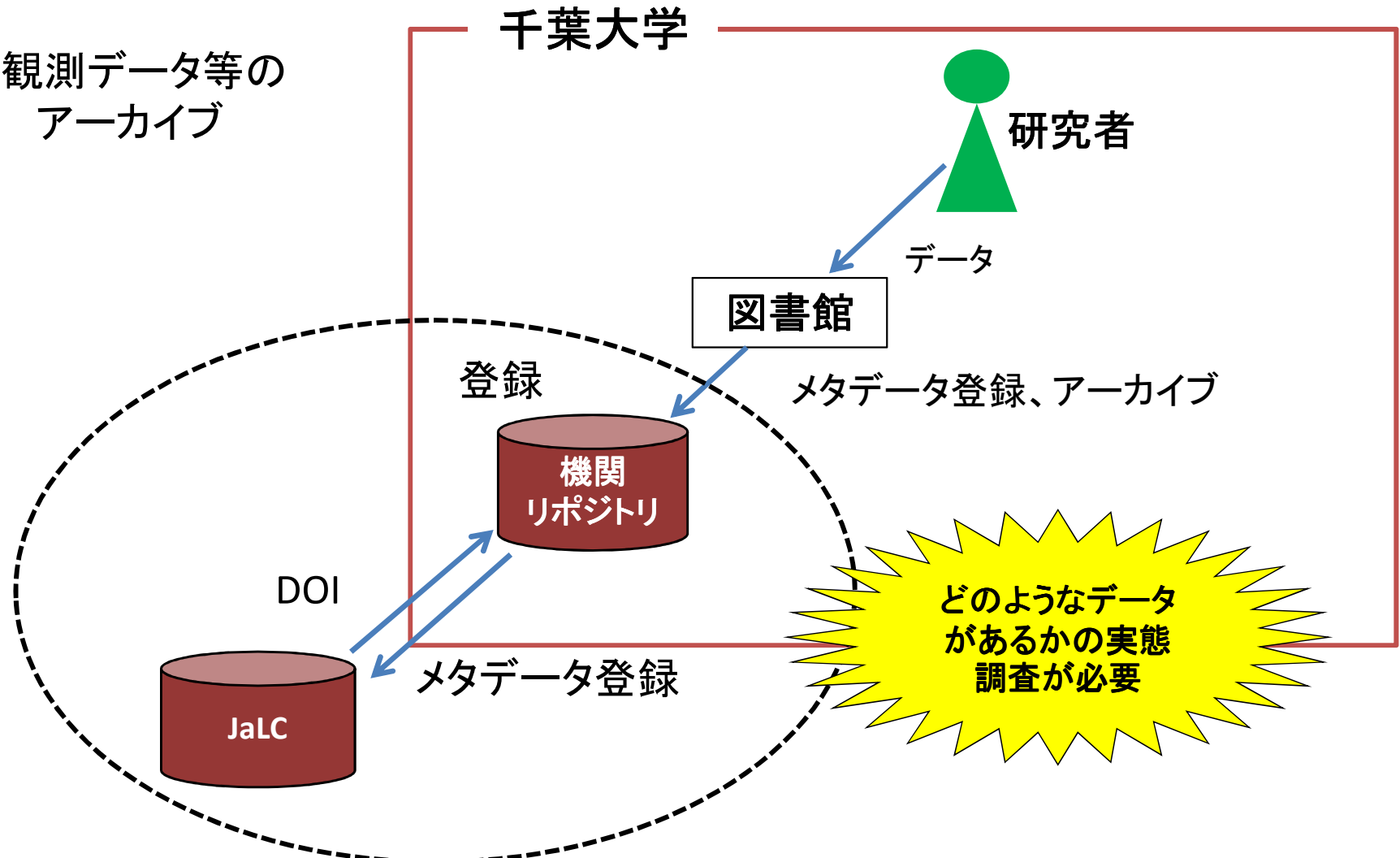
## 2) 外部資金の助成による研究成果の登録フロー

- ・学内の研究推進部門と連携した運営フローを想定。
- ・助成機関のOAポリシーや研究不正対応のガイドラインも考慮



# 課題：業務フロー検討案

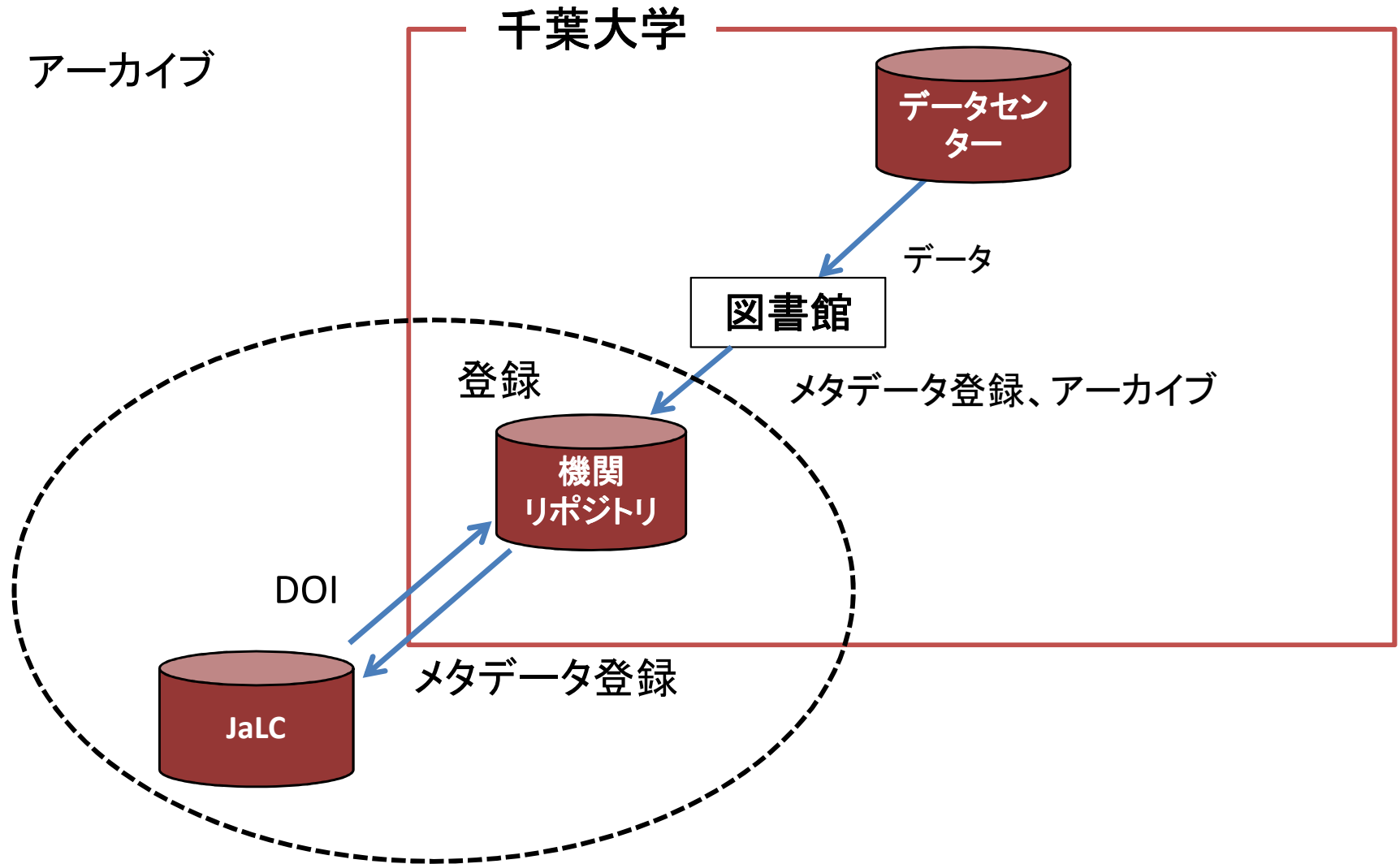
## 3) 観測データ等の アーカイブ



研究データそのものを機関リポジトリで  
アーカイブすることは現実的ではない  
かもしれないが、機関として

# 課題：業務フロー検討案

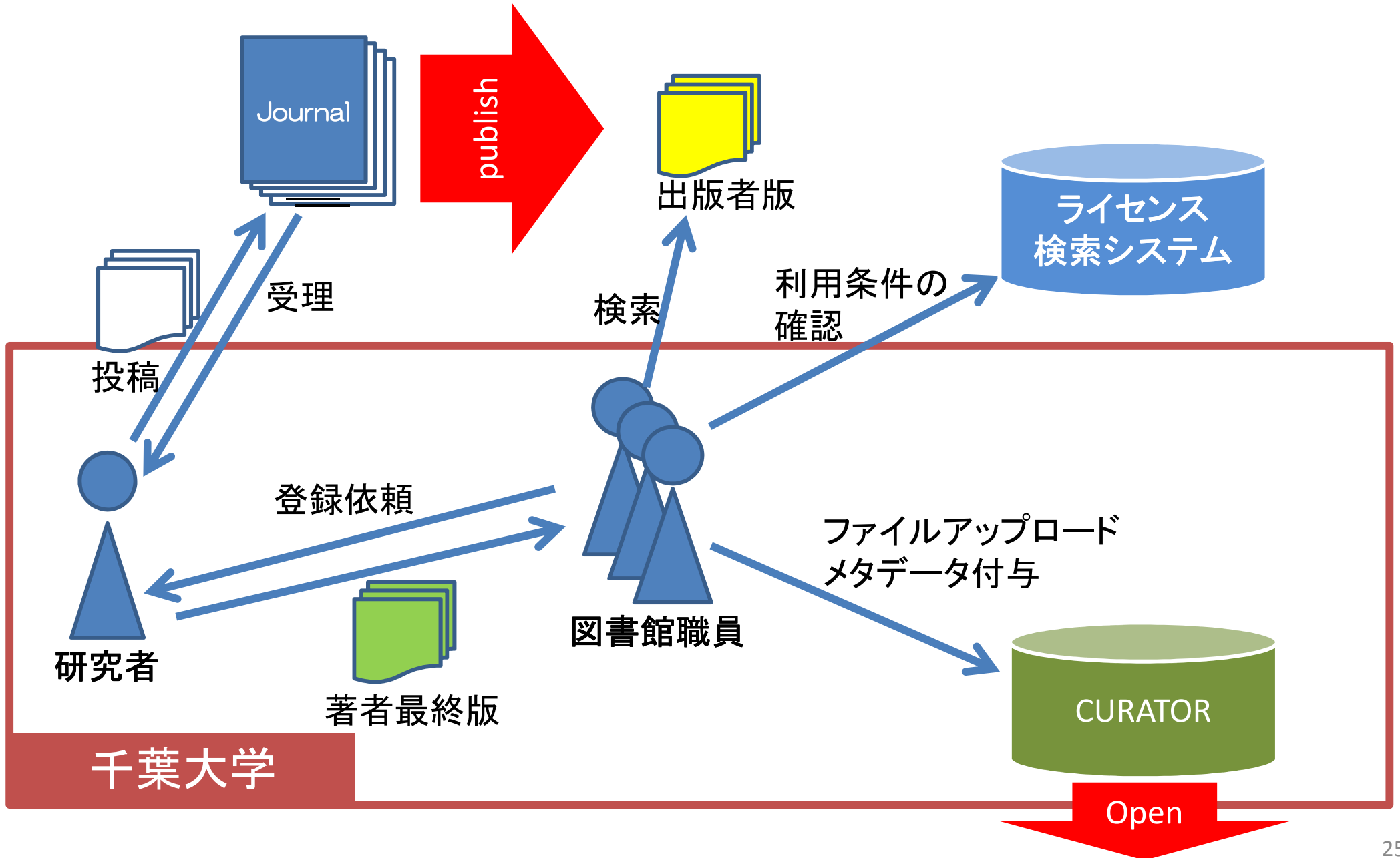
## 3) アーカイブ



研究データそのものを機関リポジトリでアーカイブすることは現実的ではないかもしれないが、**機関としてリソースのカタログを作りDOIを付与することは重要ではないか？**



# 機関リポジトリ登録フロー(図書館依頼)



# 作業負荷

- 実はメタデータ付与、ファイルアップロードよりも、**ライセンス処理**や**研究者との連絡調整の前処理**の負荷のほうがはるかに重い(研究データの著作権は?)。
- メタデータ付与やファイルアップロードは、手順が整理できれば、それほど大変ではない。
- 本来は、研究者が自ら発信するものだが、待っていたら集まらないので、現実には図書館員が研究者に個別にコンタクトして、**登録発信を促している**例が多い。
- **研究業績データベース**との連携は一部の大学ではリンク機能を実装済み。ただし**標準ができていない**(機関リポジトリ側には著者IDが入っていない、業績データベース側はそもそもメタデータ標準がない)ので、**システム継承は力技**(ResearchMapの利用は拡大するか?)。

# 課題：データマネジメントのポリシー

---

日本の大学図書館で作成しているところは、おそらくまだないが、何を集めて何を集めないのか？データ移行をどう考えるか？データ粒度をどう考えるか？などなど検討しなければならないことだらけ。

論文より明らかに複雑。そもそも**データの定義**からはじめなければならない。

個々の分野では参考例があるが、総合的に参考にする例がない。

# 課題：メタデータスキーマ

---

○機関リポジトリの国内のスキーマ:junii2

<http://www.nii.ac.jp/irp/archive/system/junii2.html>

○ JaLCのスキーマ

○ DataCite

<https://www.datacite.org/>

いずれを用いるにせよ、まともにやろうと  
思ったらシステムの改修が必要

# スキーマだけでは品質は保持できない

---

**入力(記述)マニュアルは必要！**

コンテンツのどこをどのように見て、どのように転記して、どのように転記するかは、図書館員では判断できない。

しかし、**マニュアルがあれば**、それを守るのが図書館員。

# 課題：システム

---

- Dspace、CKAN等、何を選ぶのか？
- ディスク容量はどれぐらい必要か？
- メタデータスキーマは？
- DOI付与システムは？
- 既存の機関リポジトリシステムでどこまで対応可能か？
- ワークスペースとオープンスペースとアーカイブスペースは一緒に良いのか？（机と書架と書庫）

# 大学図書館全体のとりくみ

---

## 機関リポジトリ推進委員会

<https://ir-suishin.repo.nii.ac.jp/>

「大学共同利用機関法人情報・システム研究機構国立情報学研究所と国公立大学図書館協力委員会により設立、「機関リポジトリを通じた大学の知の発信システムの構築」に関する事項を企画・立案し、学術情報の円滑な流通及び発信力の強化にかかる活動を推進することを目的とする。

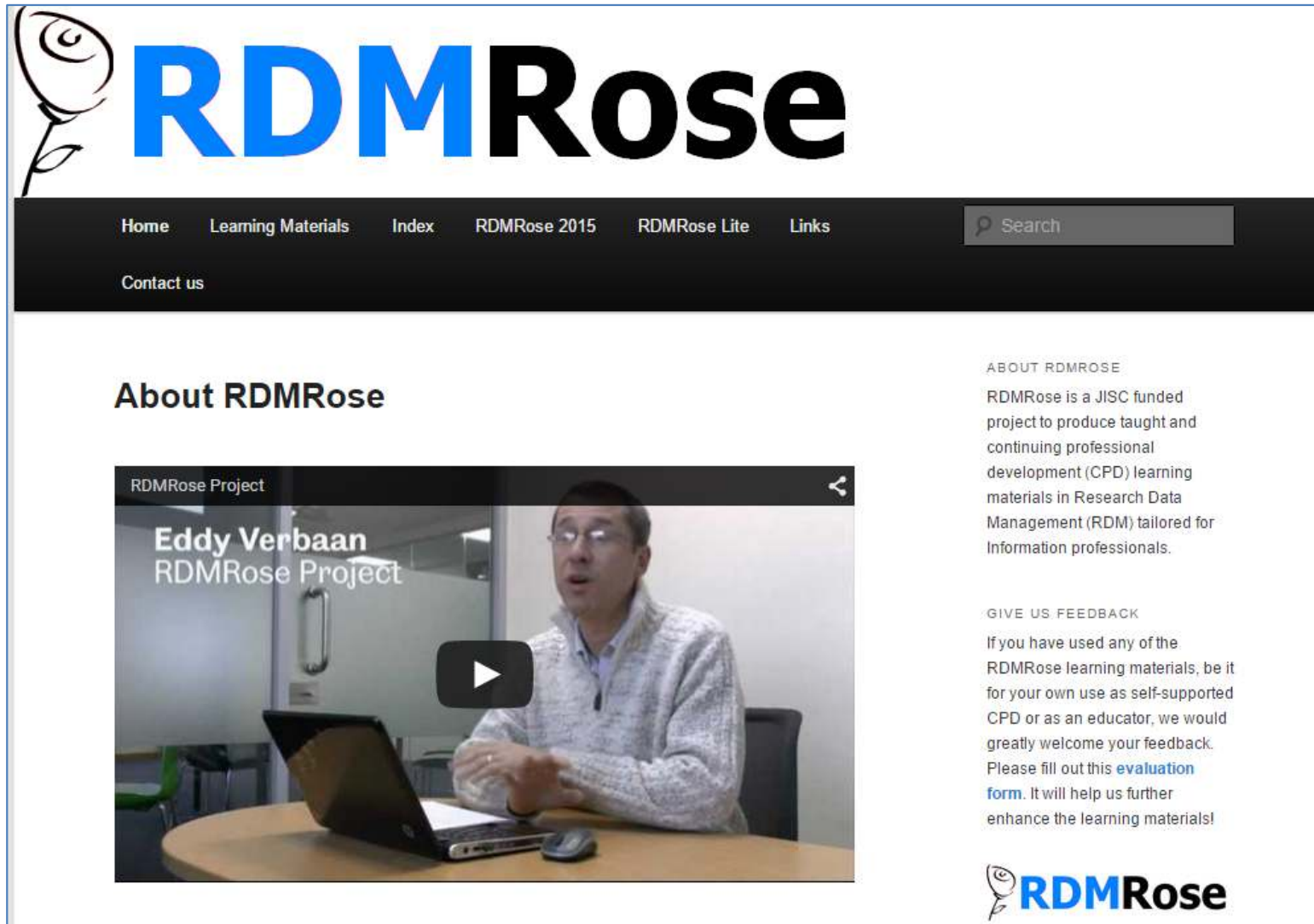
# オープンサイエンス班の平成28年度活動計画

---

- 【1】OAポリシーの策定支援ツールの開発
- 【2】RDMトレーニングツールの開発
- 【3】研究データ対応メタデータスキーマの検討
- 【4】ケーススタディによる研究データ管理  
ノウハウの蓄積



# RDMトレーニングツールの例



The screenshot shows the RDMRose website homepage. At the top left is a logo featuring a stylized rose flower next to the text "RDMRose". Below the logo is a dark navigation bar with links for "Home", "Learning Materials", "Index", "RDMRose 2015", "RDMRose Lite", and "Links". A search bar is located on the right side of the navigation bar. Below the navigation bar is a "Contact us" link. The main content area is divided into two columns. The left column has a heading "About RDMRose" and a video player. The video player shows a man, Eddy Verbaan, sitting at a desk with a laptop, speaking. The video title is "RDMRose Project" and the speaker's name is "Eddy Verbaan RDMRose Project". The right column has a heading "ABOUT RDMROSE" and a paragraph of text: "RDMRose is a JISC funded project to produce taught and continuing professional development (CPD) learning materials in Research Data Management (RDM) tailored for Information professionals." Below this is a heading "GIVE US FEEDBACK" and a paragraph of text: "If you have used any of the RDMRose learning materials, be it for your own use as self-supported CPD or as an educator, we would greatly welcome your feedback. Please fill out this [evaluation form](#). It will help us further enhance the learning materials!" At the bottom right of the page is the RDMRose logo.

**RDMRose**

Home Learning Materials Index RDMRose 2015 RDMRose Lite Links

Search

Contact us

## About RDMRose

RDMRose Project

Eddy Verbaan  
RDMRose Project

ABOUT RDMROSE

RDMRose is a JISC funded project to produce taught and continuing professional development (CPD) learning materials in Research Data Management (RDM) tailored for Information professionals.

GIVE US FEEDBACK

If you have used any of the RDMRose learning materials, be it for your own use as self-supported CPD or as an educator, we would greatly welcome your feedback. Please fill out this [evaluation form](#). It will help us further enhance the learning materials!

**RDMRose**

<http://rdmrose.group.shef.ac.uk/>

# RDMトレーニングツールの例

The screenshot shows the website for 'Essentials 4 Data Support' by RDNL (Research Data Netherlands). The header includes navigation links: Website RDNL, Home Course, Contact, RSS, and Search. The logo for RDNL is displayed, along with the text 'research data netherlands'. A blue banner on the right side of the header reads 'Essentials 4 Data Support'. Below the header is a black navigation bar with links: ABOUT THE COURSE, START THE COURSE, FORUM, and LOGIN. The main content area features a description of the course: 'Essentials 4 Data Support is an introductory course for those people who (want to) support researchers in storing, managing, archiving and sharing their research data. Essentials 4 Data Support is a product of Research Data Netherlands.' To the right of the text is an image of two stylized figures holding a globe covered in binary code. Below the main content are two columns: 'Information' and 'Next course'. The 'Information' column contains a button for 'ABOUT THE COURSE' with a sub-link 'Mission, target group & learning objectives' and a small image of hands holding a globe. The 'Next course' column contains a small image of a globe with a flag on top.

<http://datasupport.researchdata.nl/en/>

# RDMトレーニングツールの例

The screenshot shows the MANTRA website interface. At the top, there is a banner with the text "MANTRA Research Data Management Training" and a colorful abstract background. Below the banner, a description states: "MANTRA is a free online course designed for researchers or others who manage digital data as part of a research project." To the right of the description are four icons representing different user roles: Research Student, Career Researcher, Senior Academic, and Information Professional. Below the description is a navigation menu with links: Home, About, Acknowledgements, DIY Training Kit for Librarians, Feedback, and Contact Us. The main content area is titled "Learning Units: Select one to start" and features a grid of nine learning unit buttons, each with an icon and a right-pointing arrow. The units are: Research data explained, File formats & transformation, Data protection, rights & access, Data management plans, Documentation, metadata, citation, Sharing, preservation & licensing, Organising data, Storage & security, and Data handling tutorials. A "Rate MANTRA (36 Votes)" button is located in the top right of the learning units section. At the bottom of the page, there is a footer containing the EDINA logo, social media icons (Twitter, Facebook, LinkedIn, Email), a Creative Commons BY license icon, and the University of Edinburgh logo. Below the logos, there is text: "EDINA Privacy & Cookies | Website Accessibility | Last update: 10 Sept 2014" and "The University of Edinburgh is a charitable body, registered in Scotland, with registration number SC005336, VAT Registration Number GB 592 9507 00."

<http://datalib.edina.ac.uk/mantra/>

*Thank You*